Stance Detection

Peter Carragher

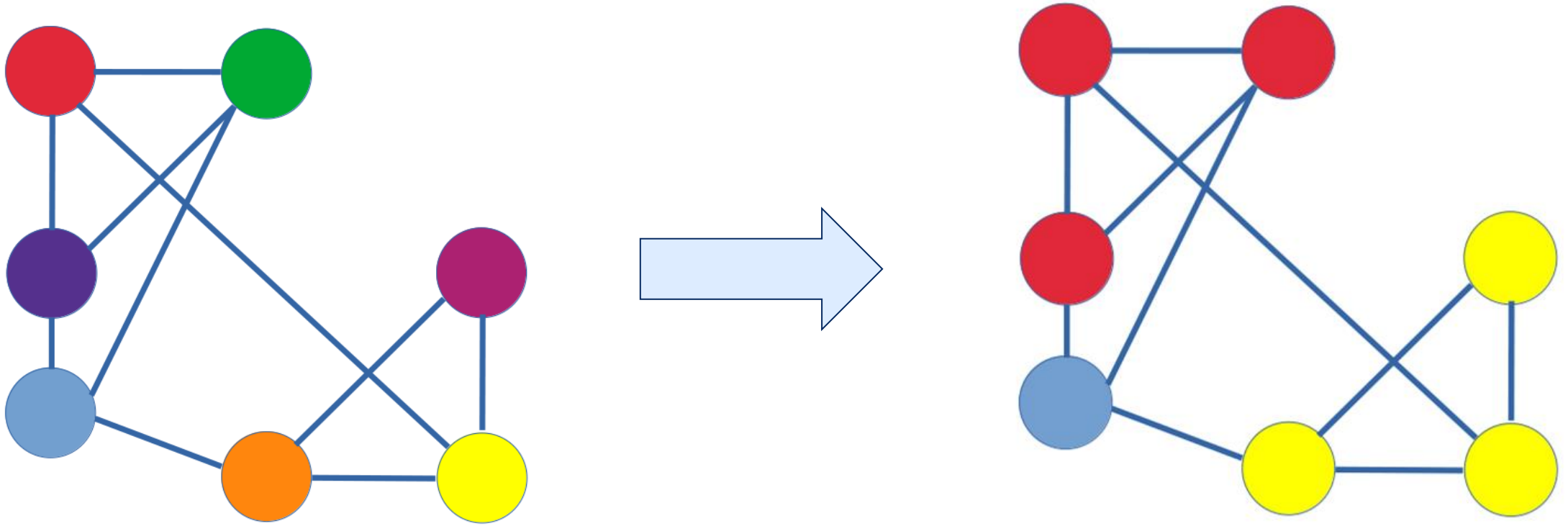
The CMU centers for:

Informed DEmocracy And Social cyber-security

Computational Analysis of Social and Organizational Systems

Carnegie Mellon University

Recap: community detection = topic-oriented groups



Learning goals: ideological groups

This session will focus on characterizing the stances of users on a specific topic as pro, anti, or neutral.

Learn the process of labeling the viewpoint or position of an agent with respect to this topic (i.e. a political issue).

Start by labelling seed hashtags that are clearly ideological and popular as pro, con, and neutral.

Example uses:

- Track popularity of certain stance over time
- See which users are pushing a position
- Find pockets of users with one stance or another

Stance detection

We will cover two stance detection algorithms:

1. Older stance propagation system developed by Sumeet Kumar [1]
   - Also known as “Document data label propagation”
2. Newer weighted stance propagation developed by Evan Williams [2]
   - Also known as “General label propagation”

Document data label propagation

Document data label propagation

Inputs:

- Agent-by-agent interaction network
- Agent-by-concept usage network
- Concepts (hashtags, URLs, etc.) labeled with stances ∈ {-1, 0, 1}

Optional inputs:

- Agent-by-document authorship network
- Document-by-word usage network

Outputs:

- Agent stance ∈ {-1, 0, 1} and confidence ∈ [0, 1]
- Concept stance ∈ {-1, 0, 1} and confidence ∈ [0, 1]

Document data label propagation

First, use labeled hashtags to label stance of users who used those hashtags

Then, repeat steps:

1. Spread stance from labeled users to unlabeled hashtags and users
2. Spread stance from labeled hashtags and users to unlabeled users

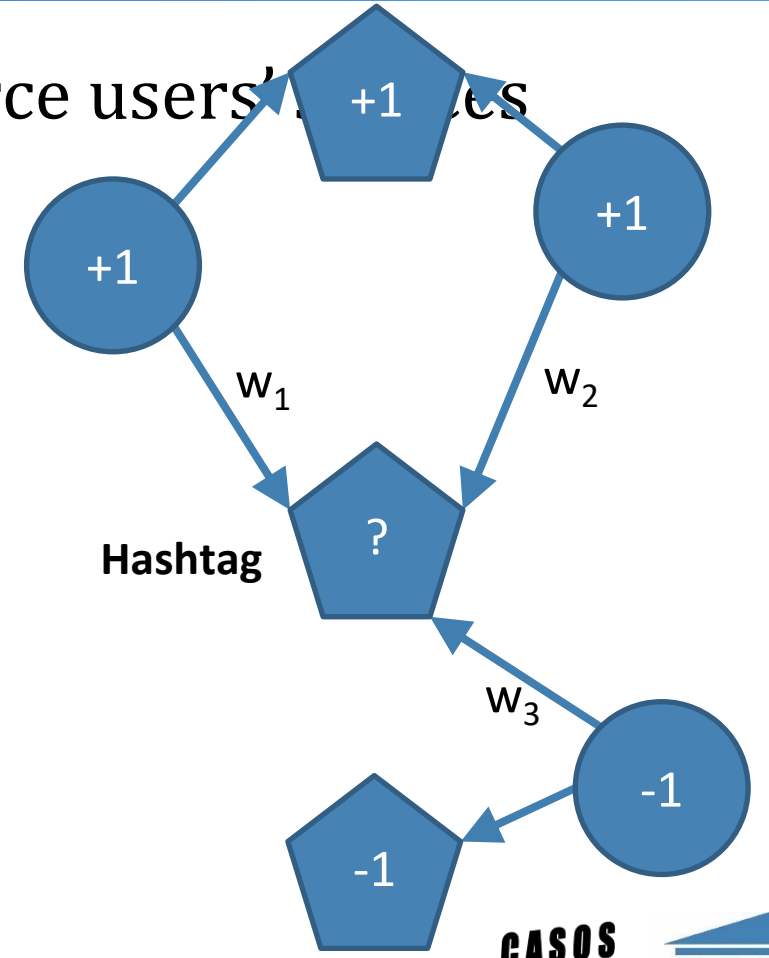
1. Spread stance from users to hashtags

Sum up usage weights multiplied by their source users' [illegible]es

- Example: $w_1 + w_2 - w_3$

If sum > threshold, assign stance +1

If sum < –threshold, assign stance -1



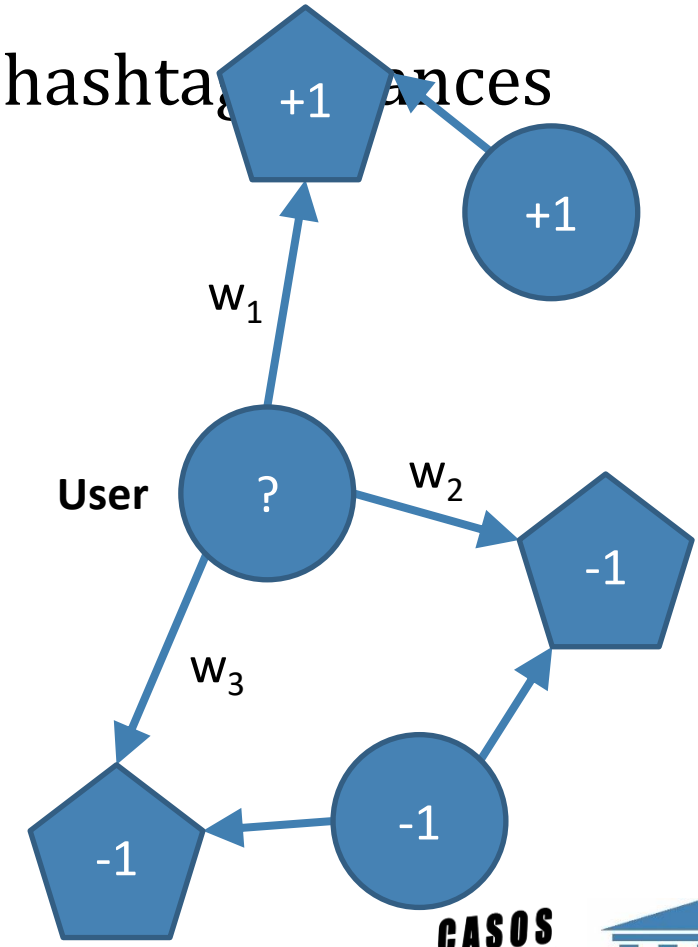
2. Spread stance from hashtags to users

Sum up usage weights multiplied by their target hashtag stances

- Example: $w_1 - w_2 - w_3$

If sum > threshold, assign stance +1

If sum < –threshold, assign stance -1



Caveats

In each step, stance is also propagated directly from user to user

- Imagine replacing “hashtags” with “hashtags and other users” in the preceding slides

Concepts beyond hashtags can also be used, such as URLs

Optional stance detection with texts

Get initial set of labeled users

Repeat steps:

1. Run user-hashtag propagation algorithm
2. Train support vector machine (SVM) [3] on tweets from labeled users
3. Run SVM on tweets from unlabeled users
4. Merge labels from Steps 1 & 3 to update set of labeled users

General label propagation

General label propagation

Propagation-based like previous algorithm, but:

- Propagates directly between agents
- Propagates through retweet network first and then through other networks
- Weights neighbors' contributions based on similarity to the unlabeled agent

Calculating similarity between users:

1. Tokenize document texts
2. Gather user-by-term frequency matrix
3. Compute cosine similarity between the term frequency vectors of each pair of users

General label propagation

1. Multiply edge weights by user similarities

   Example: w_1s_1, w_2s_2, w_3s_3

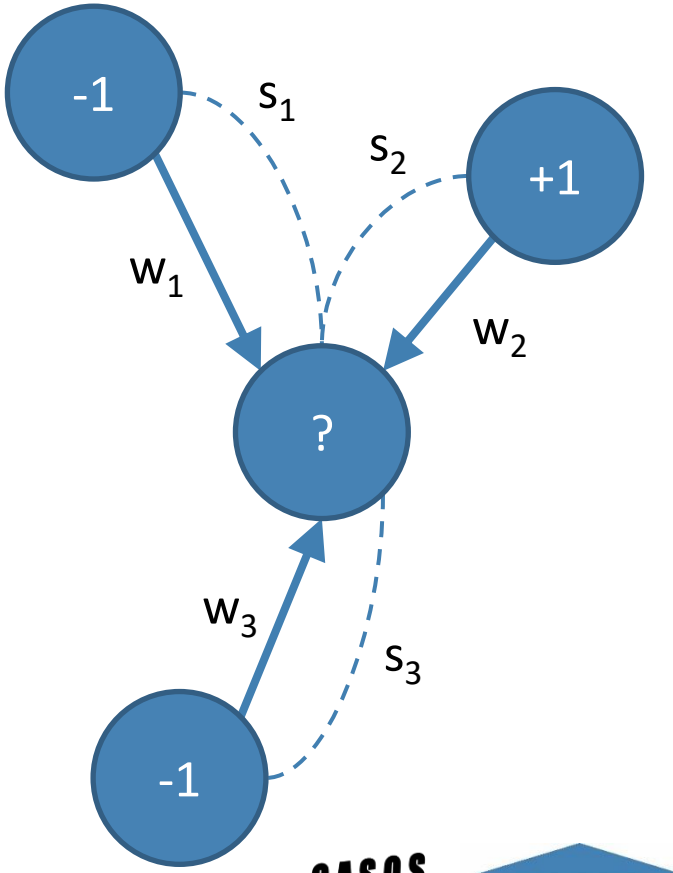
2. Sum the results for the same stance

   Pro: w_2s_2 Con: $w_1s_1 + w_3s_3$

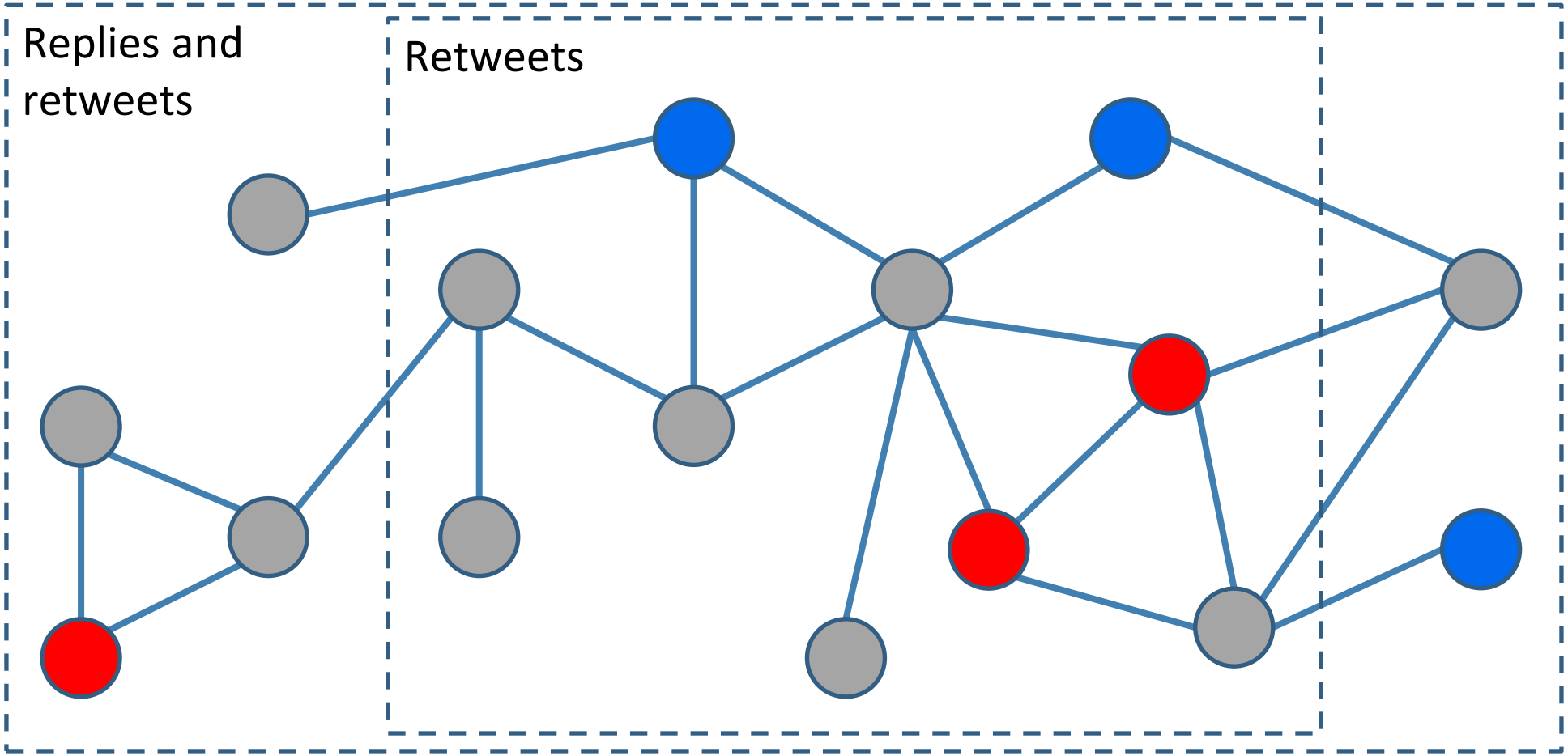
3. Take the maximum

   If $w_1s_1 + w_3s_3 > w_2s_2$, then assign con

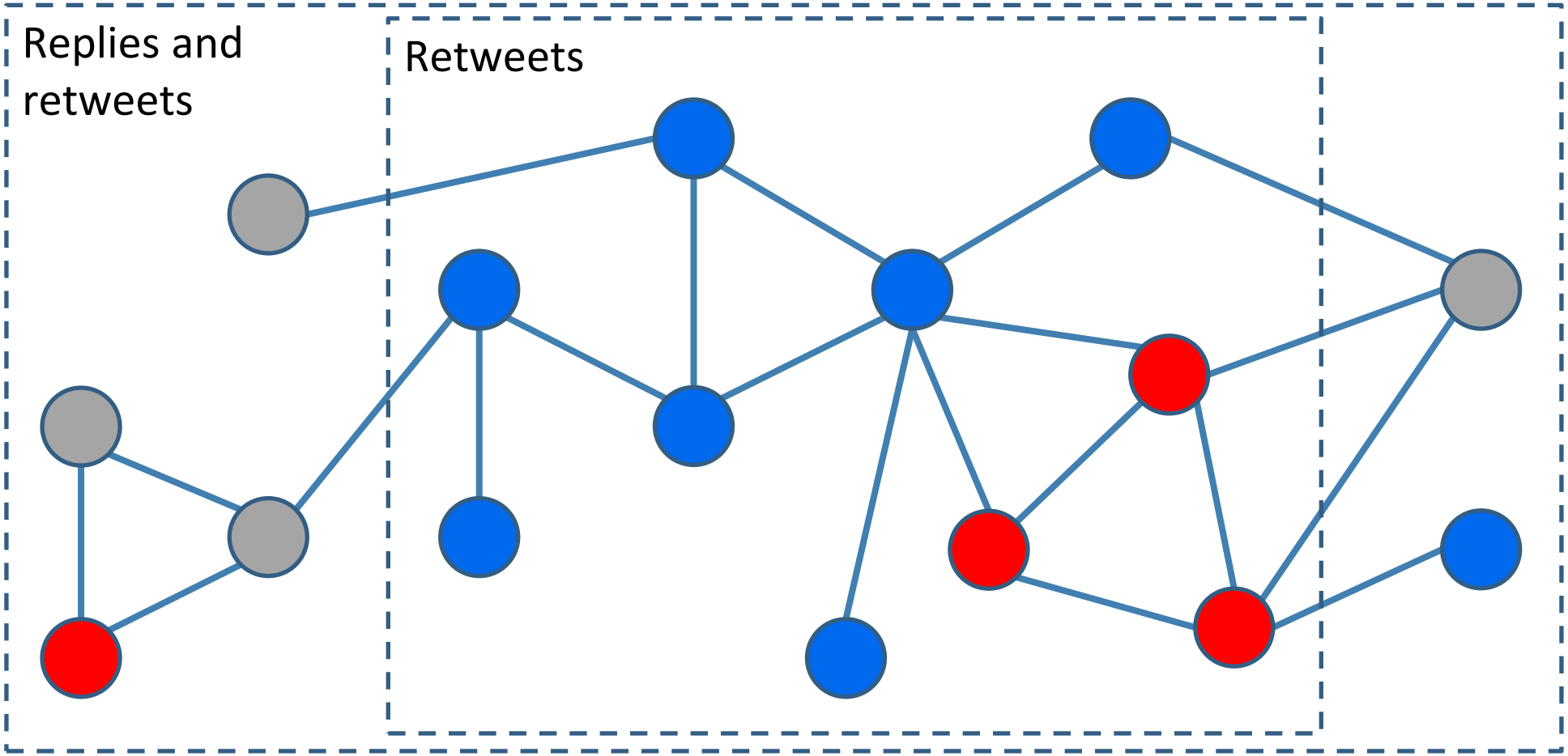
   If $w_1s_1 + w_3s_3 < w_2s_2$, then assign pro



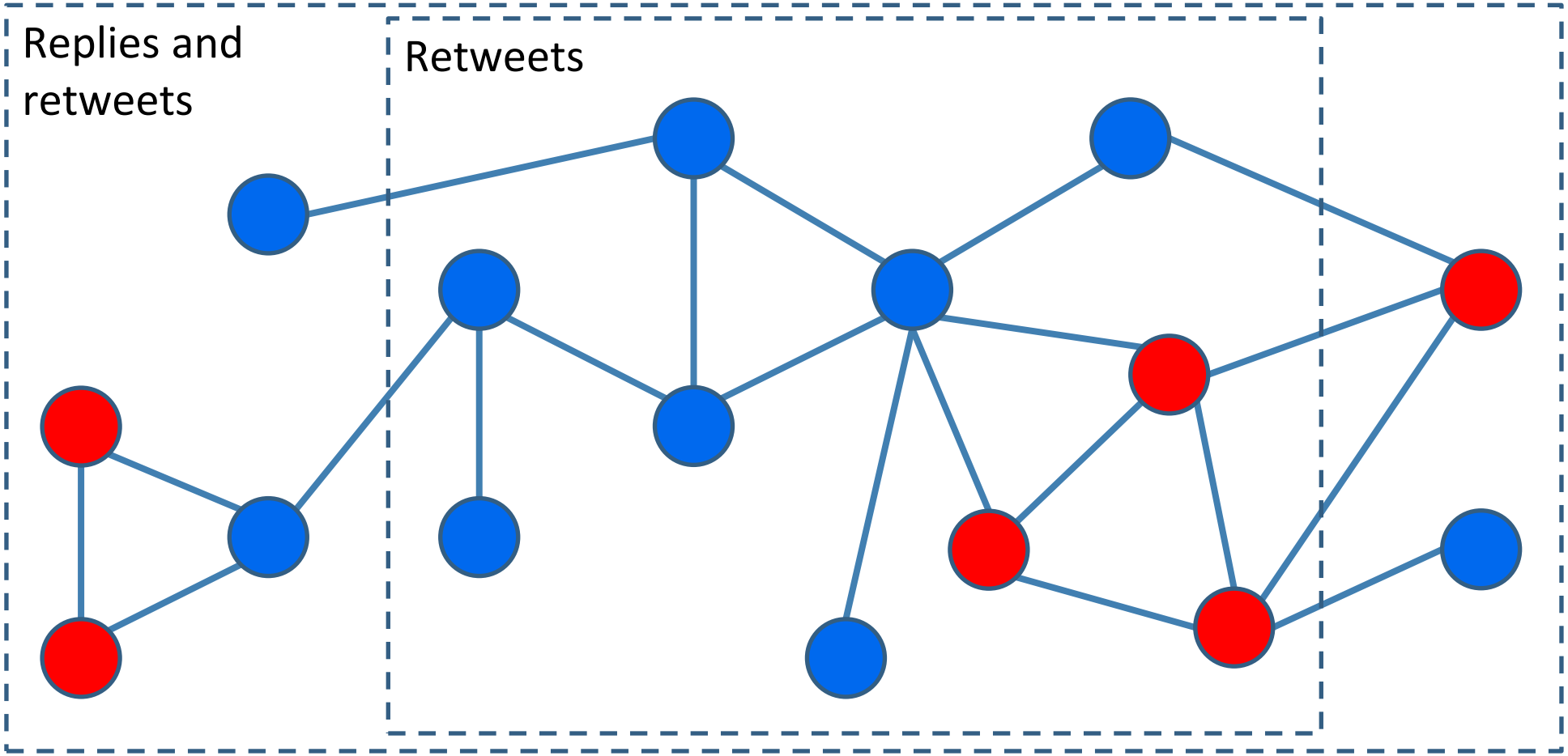
1. Use hashtags to label initial users



2. Propagate until convergence over retweet network



3. Propagate until convergence over larger network



Which algorithm should I use?

Algorithm	Advantages	Disadvantages
Document data label propagation	• Stance can better spread to disconnected components via hashtags/concepts • Can provide explicit document-by-word network rather than relying on built-in tokenization	• Slow when providing document texts • Relies on a threshold when determining stance
General label propagation	• Fast, particularly if providing document texts • Does not rely on a threshold	• Convergence to single solution not guaranteed • May be less effective for smaller components of network

References

1. Kumar, S., Villa Cox, R., Babcock, M., & Carley, K. M. (2017). “A Weakly Supervised Approach for Classifying Stance in Twitter Replies”. *Arxiv*: 2103.07098.

2. Williams, E., & Carley, K. M. (2022). “TSPA: Efficient Target-Stance Detection on Twitter”. *2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*: 242–246. doi: 10.1109/ASONAM55673.2022.10068608

3. Bottou, L. (2010). “SVM with Averaged Stochastic Gradient Descent (ASGD)”. https://leon.bottou.org/projects/sgd